3

Research designs

Chapter outline

Introduction 45

Criteria in social research 46

- Reliability 46
- Replication 47
- Validity 47
- Relationship with research strategy 48

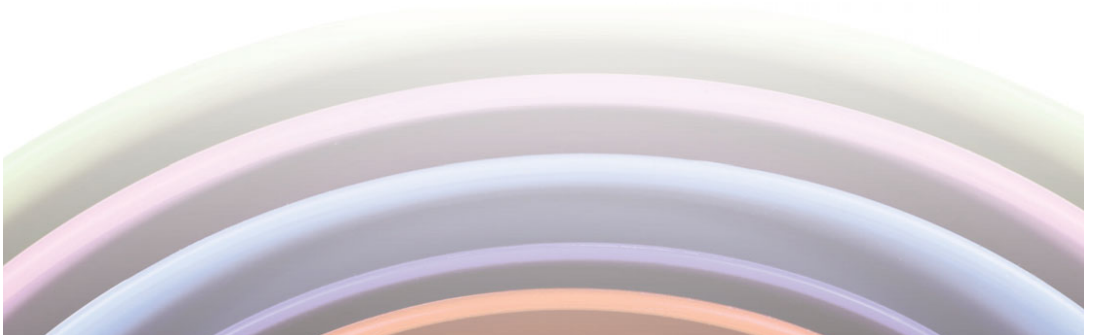
Research designs 50

- Experimental design 50
- Cross-sectional design 59
- Longitudinal design(s) 63
- Case study design 66
- Comparative design 72

Bringing research strategy and research design together 76

Key points 77

Questions for review 77



Chapter guide

In focusing on the different kinds of research design, we are paying attention to the different frameworks for the collection and analysis of data. A research design relates to the criteria that are employed when evaluating social research. It is, therefore, a framework for the generation of evidence that is suited both to a certain set of criteria and to the research question in which the investigator is interested. This chapter is structured as follows.

- Reliability, replication, and validity are presented as criteria for assessing the quality of social research. The latter entails an assessment in terms of several criteria covered in the chapter: measurement validity; internal validity; external validity; and ecological validity.
- The suggestion that such criteria are mainly relevant to quantitative research is examined, along with the proposition that an alternative set of criteria should be employed in relation to qualitative research. This alternative set of criteria, which is concerned with the issue of trustworthiness, is outlined briefly.
- Five prominent research designs are then outlined:
  - experimental and related designs (such as the quasi-experiment);
  - cross-sectional design, the most common form of which is survey research;
  - longitudinal design and its various forms, such as the panel study and the cohort study;
  - case study design;
  - comparative design.
- Each research design is considered in terms of the criteria for evaluating research findings.

Introduction

In the previous chapter, the idea of *research strategy* was introduced as a broad orientation to social research. The specific context for its introduction was the distinction between quantitative and qualitative research as different research strategies. However, the decision to adopt one or the other strategy will not get you far along the road of doing a piece of research. Two other key decisions will have to be made (along with a host of tactical decisions about the way in which the research will be carried out and the data analysed). These decisions concern choices about *research design* and *research method*. On the face of it, these two terms would seem to mean the same thing, but it is crucial to draw a distinction between them (see Key concepts 3.1 and 3.2).

Research methods can be and are associated with different kinds of research design. The latter represents a structure that guides the execution of a research method and the analysis of the subsequent data. The two terms are often confused. For example, one of the research designs to be covered in this chapter—the case study—is very often referred to as a *method*. As we will see, a case study entails the detailed exploration of a specific case, which could be a community, organization, or person. But, once a case has been selected, a research method or research methods are needed to collect data. Simply selecting an organization and deciding to study it intensively are not going to provide data. Do you observe? Do you conduct interviews? Do you examine documents? Do you administer questionnaires? You may in fact use any or all of these research methods, but the crucial point is that choosing a case study approach will not in its own right provide you with data.

Key concept 3.1
What is a research design?

A **research design** provides a framework for the collection and analysis of data. A choice of research design reflects decisions about the priority being given to a range of dimensions of the research process. These include the importance attached to:

- expressing causal connections between **variables**;
- generalizing to larger groups of individuals than those actually forming part of the investigation;
- understanding behaviour and the meaning of that behaviour in its specific social context;
- having a temporal (that is, over time) appreciation of social phenomena and their interconnections.

Key concept 3.2
What is a research method?

A research method is simply a technique for collecting data. It can involve a specific instrument, such as a self-completion questionnaire or a structured **interview schedule**, or participant observation whereby the researcher listens to and watches others.

In this chapter, five different research designs will be examined: experimental design and its variants, including **quasi-experiments**; cross-sectional or survey design; longitudinal design; case study design; and **comparative design**. However, before embarking on the nature of and differences between these designs, it is useful to consider some recurring issues in social research that cut across some or all of these designs.

Criteria in social research

Three of the most prominent criteria for the evaluation of social research are **reliability**, **replication**, and **validity**. Each of these terms will be treated in much greater detail in later chapters, but in the meantime a fairly basic treatment of them can be helpful.

Reliability

Reliability is concerned with the question of whether the results of a study are repeatable. The term is commonly used in relation to the question of whether the measures that are devised for concepts in the social sciences (such as poverty, racial prejudice, deskilling, religious orthodoxy) are consistent. In Chapter 7 we will be looking at the idea of reliability in greater detail, in particular the different ways in which it can be conceptualized. Reliability is particularly at issue in connection with quantitative research. The quantitative researcher is likely to be concerned with the question of whether a measure is stable or not. After all, if we found that IQ tests, which were designed as measures of intelligence, were found to fluctuate, so that people's IQ scores were often wildly different when administered on two or more occasions, we would be concerned about it as a measure. We would consider it an unreliable measure—we could not have faith in its consistency.

Replication

The idea of reliability is very close to another criterion of research—replication and more especially replicability. It sometimes happens that researchers choose to replicate the findings of others. There may be a host of different reasons for doing so, such as a feeling that the original results do not match other evidence that is relevant to the domain in question. In order for replication to take place, a study must be capable of replication—it must be replic*able*. This is a very obvious point: if a researcher does not spell out his or her procedures in great detail, replication is impossible. Similarly, in order for us to assess the reliability of a measure of a concept, the procedures that constitute that measure must be replicable by someone else. Ironically, replication in social research is not common. In fact, it is probably truer to say that it is quite rare. When Burawoy (1979) found that by accident he was conducting case study research in a US factory that had been studied three decades earlier by another researcher (Donald Roy), he thought about treating his own investigation as a replication. However, the low status of replication in academic life persuaded him to resist this option. He writes: 'I knew that to replicate Roy's study would not earn me a dissertation let alone a job. . . . [In] academia the real reward comes not from replication but from originality!' (Burawoy 2003: 650). Nonetheless, an investigation's capacity to be replicated—replicability—is highly valued by many social researchers working within a quantitative research tradition. See Research in focus 7.7 for an example of a replication study.

Validity

A further and in many ways the most important criterion of research is *validity*. Validity is concerned with the integrity of the conclusions that are generated from a piece of research. As we shall do for reliability, we will be examining the idea of validity in greater detail in later chapters, but in the meantime it is important to be aware of the main types of validity that are typically distinguished:

- *Measurement validity*. **Measurement validity** applies primarily to quantitative research and to the search for measures of social scientific concepts. Measurement validity is also often referred to as *construct validity*. Essentially, it is to do with the question of whether a measure that is devised of a concept really does reflect the concept that it is supposed to be denoting. Does the IQ test really measure variations in intelligence? If we take the study reported in Research in focus 2.4, there are three concepts that needed to be measured in order to test the hypotheses: national religiosity, religious orthodoxy, and family religious orientation. The question then is: do the measures really represent the concepts they are supposed to be tapping? If they do not, the study's findings will be questionable. It should be appreciated that measurement validity is related to reliability: if a measure of a concept is unstable in that it fluctuates and hence is unreliable, it simply cannot be providing a valid measure of the concept in question. In other words, the assessment of measurement validity presupposes that a measure is reliable. If a measure is unreliable because it does not give a stable reading of the underlying concept, it cannot be valid, because a valid measure reflects the concept it is supposed to be measuring.
- *Internal validity*. **Internal validity** relates mainly to the issue of **causality**, which will be dealt with in greater detail in Chapter 7. Internal validity is concerned with the question of whether a conclusion that incorporates a causal relationship between two or more variables holds water. If we suggest that *x* causes *y*, can we be sure that it is *x* that is responsible for variation in *y* and not something else that is producing an apparent causal relationship? In the study examined in Research in focus 2.4, the authors were quoted as concluding that 'the religious environment of a nation has a major impact on the beliefs of its citizens' (Kelley and De Graaf 1997: 654). Internal validity raises the question: can we be sure that national religiosity really does cause variation in religious orientation and that this apparent causal relationship is genuine and not produced by something else? In discussing issues of causality, it is common to refer to the factor that has a causal impact as the **independent variable** and the effect as the **dependent variable** (see Key concept 3.3). In the case of Kelley and De Graaf's research, the 'religious environment of a nation' was an independent variable and 'religious belief' was the dependent variable. Thus, internal validity raises the question: how confident can we be that the independent variable really is at least in part responsible for the variation that has been identified in the dependent variable?
- *External validity*. **External validity** is concerned with the question of whether the results of a study can be generalized beyond the specific research context. In the research by Poortinga et al. (2004) on foot and mouth disease that was referred to in Research in focus 2.8, data were collected from 229 respondents in Bude and 244 respondents in Norwich. Can their findings

Key concept 3.3
What is a variable?

A variable is simply an attribute on which cases vary. 'Cases' can obviously be people, but they can also include things such as households, cities, organizations, schools, and nations. If an attribute does not vary, it is a **constant**. If all manufacturing organizations had the same ratio of male to female managers, this attribute of such organizations would be a constant and not a variable. Constants are rarely of interest to social researchers. It is common to distinguish between different types of variable. The most basic distinction is between *independent variables* and *dependent variables*. The former are deemed to have a causal influence on the latter. In addition, it is important to distinguish between variables—whether independent or dependent—in terms of their measurement properties. This is an important issue in the context of quantitative data analysis. In Chapter 15, a distinction is drawn between the following types of variable: **interval/ratio variables**; **ordinal variables**; **nominal variables**; and **dichotomous variables**. See page 335 for an explanation of these main types and Table 15.1 for brief descriptions of them.

about the attitudes to the handling of the outbreak be generalized beyond these respondents? In other words, if the research was not externally valid, it would apply to the 473 respondents alone. If it was externally valid, we would expect it to apply more generally to the populations of these two towns at the time of the outbreak of the disease. It is in this context that the issue of how people are selected to participate in research becomes crucial. This is one of the main reasons why quantitative researchers are so keen to generate representative samples (see Chapter 8).

- *Ecological validity*. **Ecological validity** is concerned with the question of whether social scientific findings are applicable to people's everyday, natural social settings. As Cicourel (1982: 15) has put it: 'Do our instruments capture the daily life conditions, opinions, values, attitudes, and knowledge base of those we study as expressed in their natural habitat?' This criterion is concerned with the question of whether social research sometimes produces findings that may be technically valid but have little to do with what happens in people's everyday lives. If research findings are ecologically *in*valid, they are in a sense *artefacts* of the social scientist's arsenal of data collection and analytic tools. The more the social scientist intervenes in natural settings or creates unnatural ones, such as a laboratory or even a special room to carry out interviews, the more likely it is that findings will be ecologically invalid. The findings deriving from a study using questionnaires may have measurement validity and a reasonable level of internal validity, and they may be externally valid, in the sense that they can be generalized to other samples confronted by the same questionnaire, but the unnaturalness of the fact of having to answer a questionnaire may mean that the findings have limited ecological validity.

Relationship with research strategy

One feature that is striking about most of the discussion so far is that it seems to be geared mainly to quantitative rather than to qualitative research. Both reliability and measurement validity are essentially concerned with the adequacy of measures, which are most obviously a concern in quantitative research. Internal validity is concerned with the soundness of findings that specify a causal connection, an issue that is most commonly of concern to quantitative researchers. External validity *may* be relevant to qualitative research, but the whole question of representativeness of research subjects with which the issue is concerned has a more obvious application to the realm of quantitative research, with its preoccupation with sampling procedures that maximize the opportunity for generating a representative sample. The issue of ecological validity relates to the naturalness of the research approach and seems to have considerable relevance to both qualitative and quantitative research.

Some writers have sought to apply the concepts of reliability and validity to the practice of qualitative research (e.g. LeCompte and Goetz 1982; Kirk and Miller 1986; Peräkylä 1997), but others argue that the grounding of these ideas in quantitative research renders them inapplicable to or inappropriate for qualitative research. Writers like Kirk and Miller (1986) have applied concepts of validity and reliability to qualitative research but have changed the sense in which the terms are used very slightly. Some qualitative researchers sometimes propose that the studies they produce should be judged or

evaluated according to different criteria from those used in relation to quantitative research. Lincoln and Guba (1985) propose that alternative terms and ways of assessing qualitative research are required. For example, they propose **trustworthiness** as a criterion of how good a qualitative study is. Each aspect of trustworthiness has a parallel with the quantitative research criteria.

- *Credibility*, which parallels internal validity—that is, how believable are the findings?
- *Transferability*, which parallels external validity—that is, do the findings apply to other contexts?
- *Dependability*, which parallels reliability—that is, are the findings likely to apply at other times?
- *Confirmability*, which parallels objectivity—that is, has the investigator allowed his or her values to intrude to a high degree?

These criteria will be returned to in Chapter 17.

Hammersley (1992*a*) occupies a kind of middle position here, in that, while he proposes validity as an important criterion (in the sense that an empirical account must be plausible and credible and should take into account the amount and kind of evidence used in relation to an account), he also proposes *relevance* as a criterion. Relevance is taken to be assessed from the vantage point of the importance of a topic within its substantive field or the contribution it makes to the literature on that field. The issues in these different views have to do with the

Key concept 3.4
What is naturalism?

Naturalism is an interesting example of a mercifully rare instance of a term that not only has different meanings, but also has meanings that can actually be contradictory! It is possible to identify three different meanings.

- *Naturalism means viewing all objects of study—whether natural or social ones—as belonging to the same realm and a consequent commitment to the principles of natural scientific method*. This meaning, which has clear affinities with positivism, implies that all entities belong to the same order of things, so that there is no essential difference between the objects of the natural sciences and those of the social sciences (M. Williams 2000). For many naturalists, this principle implies that there should be no difference between the natural and the social sciences in the ways in which they study phenomena. This version of naturalism essentially proposes that there is a unity between the objects of the natural and the social sciences and that, because of this, there is no reason for social scientists not to employ the approaches of the natural scientist.
- *Naturalism means being true to the nature of the phenomenon being investigated*. According to Matza, naturalism is 'the philosophical view that strives to remain true to the nature of the phenomenon under study' (1969: 5) and 'claims fidelity to the natural world' (1969: 8). This meaning of the term represents a fusion of elements of an interpretivist epistemology and a constructionist ontology, which were examined in Chapter 2. Naturalism is taken to recognize that people attribute meaning to behaviour and are authors of their social world rather than passive objects.
- *Naturalism is a style of research that seeks to minimize the intrusion of artificial methods of data collection*. This meaning implies that the social world should be as undisturbed as possible when it is being studied (Hammersley and Atkinson 1995: 6).

The second and third meanings overlap considerably, in that it could easily be imagined that, in order to conduct a naturalistic enquiry in the second sense, a research approach that adopted naturalistic principles in the third sense would be required. Both the second and third meanings are incompatible with, and indeed opposed to, the first meaning. Naturalism in the first sense is invariably viewed by writers drawing on an interpretivist epistemology as not 'true' to the social world, precisely because: it posits that there are no differences between humans and the objects of the natural sciences; it therefore ignores the capacity of humans to interpret the social world and to be active agents; and, in its preference for the application of natural science methods, it employs artificial methods of data collection. When writers are described as *anti-naturalists*, it is invariably the first of the three meanings that they are deemed to be railing against.

different objectives that many qualitative researchers argue are distinctive about their craft. The distinctive features of qualitative research will be examined in later chapters.

However, it should also be borne in mind that one of the criteria previously cited—ecological validity—may have been formulated largely in the context of quantitative research, but is in fact a feature in relation to which qualitative research fares rather well. Qualitative research often involves a *naturalistic* stance (see Key concept 3.4). This means that the researcher seeks to collect data in naturally occurring situations and environments, as opposed to fabricated, artificial ones. This characteristic probably applies particularly well to ethnographic research, in which participant observation is a prominent element of data collection, but it is sometimes suggested that it applies also to the sort of interview approach typically used by qualitative researchers, which is less directive than the kind used in quantitative research (see e.g. Research in focus 2.4). We might expect that much qualitative research is stronger than quantitative investigations in terms of ecological validity.

By and large, these issues in social research have been presented because some of them will emerge in the context of the discussion of research designs in the next section, but in a number of ways they also represent background considerations for some of the issues to be examined. They will be returned to later in the book.

Research designs

In this discussion of research designs, five different types will be examined: experimental design; cross-sectional or survey design; longitudinal design; case study design; and comparative design. Variations on these designs will be examined in their relevant subsections.

Experimental design

True **experiments** are quite unusual in sociology, but are employed in related areas of enquiry, such as social psychology and organization studies, while researchers in social policy sometimes use them in order to assess the impact of new reforms or policies. Why, then, bother to introduce experimental designs at all in the context of a book about social research? The chief reason, quite aside from the fact that they are sometimes employed, is that a true experiment is often used as a yardstick against which non-experimental research is assessed. Experimental research is frequently held up as a touchstone because it engenders considerable confidence in the robustness and trustworthiness of causal findings. In other words, true experiments tend to be very strong in terms of internal validity.

Manipulation

If experiments are so strong in this respect, why then do social researchers not make far greater use of them? The reason is simple: in order to conduct a true experiment, it is necessary to *manipulate* the independent variable in order to determine whether it does in fact have an influence on the dependent variable. Experimental subjects are likely to be allocated to one of two or more **experimental groups**, each of which represents different types or levels of the independent variable. It is then possible to establish how far differences between the groups are responsible for variations in the level of the dependent variable. Manipulation, then, entails intervening in a situation to determine the impact of the manipulation on subjects. However, the vast majority of independent variables with which social researchers are concerned cannot be manipulated. If we are interested in the effects of gender on work experiences, we cannot *manipulate* gender so that some people are made male and others female. If we are interested in the effects of variations in social class on social and political attitudes or on health, we cannot allocate people to different social class groupings. As with the huge majority of such variables, the levels of social engineering that would be required are beyond serious contemplation.

Before moving on to a more complete discussion of experimental design, it is important to introduce a basic distinction between the *laboratory experiment* and the *field experiment*. As its name implies, the laboratory experiment takes place in a laboratory or in a contrived setting, whereas field experiments occur in real-life settings, such as in classrooms and organizations, or as a result of the implementation of reforms or new policies. It is experiments of the latter type that are most likely to touch on areas of interest to social researchers. In order to illustrate the nature of manipulation and the idea of a field experiment, Research in focus 3.1 describes a well-known piece of experimental research.

Research in focus 3.1
A field experiment

As part of a programme of research into the impact of self-fulfilling prophecies (for example, where someone's beliefs or expectations about someone else influence how the latter behaves), Rosenthal and Jacobson (1968) conducted research into the question of whether teachers' expectations of their students' abilities in fact influence the school performance of the latter. The research was conducted in a lower-class locality in the USA with a high level of children from minority group backgrounds. In the spring of 1964, all the students completed a test that was portrayed as a means of identifying 'spurters'—that is, students who were likely to excel academically. At the beginning of the following academic year, all the teachers were notified of the names of the students who had been identified as spurters. In fact, 20 per cent of the schoolchildren had been identified as spurters. However, the students had actually been administered a conventional IQ test and the so-called spurters had been selected randomly. The test was readministered eight months after the original one. The authors were then able to compare the differences between the spurters and the other students in terms of changes in various measures of academic performance, such as IQ scores, reading ability, and intellectual curiosity. Since there was no evidence for there being any difference in ability between the spurters and the rest, any indications that the spurters did in fact differ from their peers could be attributed to the fact that the teachers had been led to expect the former would perform better. The findings show that such differences did in fact exist, but that the differences between the spurters and their peers tended to be concentrated in the first two or three years of schooling. In other words, the evidence for a teacher expectancy effect was patchy. Nonetheless, this is an influential experiment that is widely believed to provide firm evidence of a teacher expectancy effect. For a useful brief review of some of the subsequent studies and reflections on Rosenthal and Jacobson's study, see Hammersley (2011: 106–9).

Classical experimental design

The research in Research in focus 3.1 includes most of the essential features of what is known as the classical experimental design, which is also often referred to as the *randomized experiment* or **randomized controlled trial** (RCT). Two groups are established, and it is this that forms the experimental manipulation and therefore the independent variable—in this case, teacher expectations. The spurters form what is known as the *experimental group* or *treatment group* and the other students form a **control group**. The experimental group receives the *experimental treatment*—teacher expectancies—but the control group does not receive an experimental treatment. The dependent variable—student performance—is measured before and after the experimental manipulation, so that a before-and-after analysis can be conducted (see Figure 3.1). Moreover, the spurters and the non-spurters were assigned randomly to their respective groups. Because of this use of **random assignment** to the experimental and control groups, the researchers were able to feel confident that the only difference between the two groups was the fact that teachers expected the spurters to fare better at school than the others. They would have been confident that, if they did establish a difference in performance between the two groups, it was due to the experimental manipulation alone.

In order to capture the essence of this design, the following simple notation will be employed:

Obs An **obs**ervation made in relation to the dependent variable; there may well be two or more observations, such as IQ test scores and reading grades before (the *pre-test*) and after (the *post-test*) the experimental manipulation.

Exp The **exp**erimental treatment (the independent variable), such as the creation of teacher expectancies. **No Exp** refers to the absence of an experimental treatment and represents the experience of the control group.

T The **t**iming of the observations made in relation to the dependent variable, such as the timing of the administration of an IQ test.

Classical experimental design and validity

What is the purpose of the control group? Surely it is what happens to the spurters (the experimental group) that really concerns us? In order for a study to be a true experiment, it must control (in other words, eliminate)

Figure 3.1

Classical experimental design (*with illustration of the effect of teacher expectancies on IQ*)

the possible effects of rival explanations of a causal finding, such as that teacher expectancies have an impact on student performance. We might then be in a position to take the view that such a study is internally valid. The presence of a control group *and* the random assignment of subjects to the experimental and control groups enable us to eliminate such rival explanations. To see this, consider some of the rival explanations that might occur if there was *no* control group. There would then have been a number of potential threats to internal validity (see Research in focus 3.2). These threats are taken from Campbell (1957) and Cook and Campbell (1979), but not all the threats to internal validity they refer to are included.

In the case of each of these threats to internal validity, each of which raises the prospect of a rival interpretation of a causal finding, the presence of a control group coupled with random assignment allows us to eliminate these threats. As a result, our confidence in the causal finding, that teacher expectancies influence student performance, is greatly enhanced.

Research in focus 3.2

Threats to internal validity (and their application to the Rosenthal and Jacobson 1968 study)

The following is a list of possible threats to the internal validity of an investigation and how each is mitigated in the Rosenthal and Jacobson (1968) study by virtue of its being a true experiment.

- *History*. This refers to events other than the manipulation of teacher expectancies that may occur in the environment and that could have caused the spurters' scores to rise. The actions of the school head to raise standards in the school may be one such type of event. If there were no control group, we could not be sure whether it was the teachers' expectancies or the head's actions that were producing the increase in spurters' grades. If there is a control group, we are able to say that history would have an effect on the control-group subjects too, and therefore differences between the experimental and control groups could be attributed to the effect of teacher expectancies alone.
- *Testing*. This threat refers to the possibility that subjects may become more experienced at taking a test or may become sensitized to the aims of the experiment as a result of the pre-test. The presence of a control group, which presumably would also experience the same effect, allows us to discount this possibility if there is a difference in levels of the dependent variable between the experimental and control groups.

- *Instrumentation*. This threat refers to the possibility that changes in the way a test is administered could account for an increase (or decrease) in scores between the pre-test and post-test—for example, if slight changes to the test had been introduced. Again, if there is a control group, we can assume that testing would have affected the control group as well.
- *Mortality*. This relates to the problem of attrition in many studies that span a long period of time, in that subjects may leave. School students may leave the area or move to a different school. Since this problem is likely to afflict the control group too, it is possible to establish its significance as a threat relative to the impact and importance of teacher expectancies.
- *Maturation*. Quite simply, people change, and the ways in which they change may have implications for the dependent variable. The students identified as spurters may have improved anyway, regardless of the effect of teacher expectancies. Maturation should affect the control group subjects as well. If we did not have a control group, it could be argued that any change in the students' school performance was attributable to the possibility that they would have improved anyway. The control group allows us to discount this possibility.
- *Selection*. If there are differences between the two groups, which would arise if they had been selected by a non-random process, variations between the experimental and control groups could be attributed to pre-existing differences in their membership. However, since a random process of assignment to the experimental and control groups was employed, this possibility can be discounted.
- *Ambiguity about the direction of causal influence*. The very notion of an independent variable and dependent variable presupposes a direction of causality. However, there may be occasions when the temporal sequence in a study is unclear, so that it is not possible to establish which variable affects the other. Since the creation of teacher expectancies preceded the improvements in academic achievement in the earlier years of school, in the Rosenthal and Jacobson study the direction of causal influence is clear.

However, simply because research is deemed to be internally valid does not mean that it is beyond reproach or that at least questions cannot be raised about it. When a quantitative research strategy has been employed, further criteria can be applied to evaluate a study. First, there is the question of measurement validity. In the case of the Rosenthal and Jacobson study, there are potentially two aspects to this. One is the question of whether academic performance has been adequately measured. Measures like reading scores seem to possess face validity, in the sense that they appear to exhibit a correspondence with what they are measuring. However, given the controversy surrounding IQ tests and what they measure (Kamin 1974), we might feel somewhat uneasy about how far gains in IQ test scores can be regarded as indicative of academic performance. Similarly, to take another of the authors' measures—intellectual curiosity—how confident can we be that this too is a valid measure of academic performance? Does it really measure what it is supposed to measure? The second question relating to measurement validity is whether the experimental manipulation really worked. In other words, did the random identification of some schoolchildren as spurters adequately create the conditions for the self-fulfilling prophecy to be examined? The procedure very much relies on the teachers being taken in by the procedure, but it is possible that they were not all equally duped. If so, this would contaminate the manipulation.

Secondly, is the research externally valid? This issue is considered in Research in focus 3.3.

Thirdly, are the findings ecologically valid? The fact that the research is a field experiment rather than a laboratory experiment seems to enhance this aspect of the Rosenthal and Jacobson research. Also, the fact that the students and the teachers seem to have had little if any appreciation of the fact that they were in fact participating in an experiment may also have enhanced ecological validity, though this aspect of the research raises enormous *ethical* concerns, since deception seems to have been a significant and probably necessary feature of the investigation. The question of ethical issues is in many ways another dimension of the validity of a study and will be the focus of Chapter 6. The fact that Rosenthal and Jacobson made intensive use of various instruments to measure academic performance might be considered a source of concerns about ecological validity, though this is an area in which most if not all quantitative research is likely to be implicated.

Research in focus 3.3

Threats to external validity (and their application to the Rosenthal and Jacobson 1968 study)

Campbell (1957) and Cook and Campbell (1979) identify five major threats to the external validity and hence **generalizability** of an investigation.

- *Interaction of selection and treatment*. This threat raises the question: to what social and psychological groups can a finding be generalized? Can it be generalized to a wide variety of individuals who might be differentiated by ethnicity, social class, region, gender, and type of personality? In the case of the Rosenthal and Jacobson study, the students were largely from lower social class groups and a large proportion were from ethnic minorities. This might be considered a limitation to the generalizability of the findings.
- *Interaction of setting and treatment*. This threat relates to the issue of how confident we can be that the results of a study can be applied to other settings. In particular, how confident can we be that Rosenthal and Jacobson's findings are generalizable to other schools? There is also the wider issue of how confident we can be that the operation of self-fulfilling prophecies can be discerned in non-educational settings. In fact, Rosenthal and others have been able to demonstrate the role and significance of the self-fulfilling prophecy in a wide variety of different contexts (Rosnow and Rosenthal 1997), though this still does not answer the question of whether the specific findings that were produced can be generalized. One set of grounds for being uneasy about Rosenthal and Jacobson's findings is that they were allowed an inordinate amount of freedom for conducting their investigation. The high level of cooperation from the school authorities was very unusual and may be indicative of the school being somewhat atypical, though whether there is any such thing as a 'typical school' is highly questionable.
- *Interaction of history and treatment*. This threat raises the question of whether the findings can be generalized to the past and to the future. The Rosenthal and Jacobson research was conducted forty years ago. How confident can we be that the findings would apply today? Also, their investigation was conducted at a particular juncture in the school academic year. Would the same results have obtained if the research had been conducted at different points in the year?
- *Interaction effects of pre-testing*. As a result of being pre-tested, subjects in an experiment may become sensitized to the experimental treatment. Consequently, the findings may not be generalizable to groups that have *not* been pre-tested, and, of course, in the real world people are rarely pre-tested in this way. The findings may therefore be partly determined by the experimental treatment as such and partly by how pre-test sensitization has influenced the way in which subjects respond to the treatment. This may have occurred in the Rosenthal and Jacobson research, since all students were pre-tested at the end of the previous academic year.
- *Reactive effects of experimental arrangements*. People are frequently, if not invariably, aware of the fact that they are participating in an experiment. Their awareness may influence how they respond to the experimental treatment and therefore affect the generalizability of the findings. Since Rosenthal and Jacobson's subjects do not appear to have been aware of the fact that they were participating in an experiment, this problem is unlikely to have been significant. The issue of **reactivity** and its potentially damaging effects is a recurring theme in relation to many methods of social research.

A fourth issue that we might want to raise relates to the question of replicability. The authors lay out very clearly the procedures and measures that they employed. If anyone were to carry out a replication, he or she would be able to obtain further information from them should they need it. Consequently, the research is replicable, although there has not been an exact replication. Clairborn (1969) conducted one of the earliest replications and followed a procedure that was very similar to Rosenthal and Jacobson's. The study was carried out in three middle-class, suburban schools, and the timing of the creation of teacher expectancies was different from that in the

original Rosenthal and Jacobson study. Clairborn failed to replicate Rosenthal and Jacobson's findings. This failure to replicate casts doubt on the external validity of the original research and suggests that the first three threats referred to in Research in focus 3.3 may have played an important part in the differences between the two sets of results.

The classical experimental design is the foundation of the randomized controlled trial, which has increasingly become the gold standard research design in health-related fields. With an RCT, the aim is to test 'alternative ways of handling a situation' (Oakley 2000: 18). This may entail comparing the impact of an intervention with what would have happened if there had been no intervention or comparing the impacts of different kinds of intervention (such as different forms of treatment of an illness). It is randomization of experimental participants that is crucial, as it means that the members of the different groups in the experiment are to all intents and purposes alike. The RCT is particularly popular in fields like medicine where research questions often take the form 'what is the impact of X?'

The laboratory experiment

Many experiments in fields like social psychology are laboratory experiments rather than field experiments. One of the main advantages of the former over the latter is that the researcher has far greater influence over the experimental arrangements. For example, it is easier to assign subjects randomly to different experimental conditions in the laboratory than to do the same in an ongoing, real-life organization. The researcher therefore has a higher level of control, and this is likely to enhance the internal validity of the study. It is also likely that laboratory experiments will be more straightforward to replicate, because they are less bound up with a certain milieu that is difficult to reproduce.

However, laboratory experiments like the one described in Research in focus 3.4 suffer from a number of limitations. First, the external validity is likely to be difficult to establish. There is the interaction of setting and treatment, since the setting of the laboratory is likely to be unrelated to real-world experiences and contexts. Also, there is likely to be an interaction of selection and treatment. In the case of Howell and Frost's (1989) study described in Research in focus 3.4, there are a number of difficulties: the subjects were students, who are unlikely to be representative of the general population, so that their responses to the experimental treatment may be distinctive; they were volunteers, and it is known that volunteers differ from non-volunteers (Rosnow and Rosenthal 1997: ch. 5); and they were given incentives to participate, which may further demarcate them from others, since not everyone is equally amenable to the blandishments of inducements. There will have been no problem of interaction effects of pre-testing, because, like many experiments, there was no pre-testing. However, it is quite feasible that reactive effects may have been set in motion by the experimental arrangements. Secondly, the ecological validity of the study may be poor, because we do not know how well the findings are applicable to the real world and everyday life. However, while the

Research in focus 3.4
A laboratory experiment

Howell and Frost (1989) were interested in the possibility that charismatic leadership, a term associated with Max Weber's (1947) types of legitimate authority, is a more effective approach to leadership in organizations than other types of leadership. They conducted a laboratory experiment that compared the effectiveness of charismatic leadership as against two other approaches—consideration and structuring. A number of hypotheses were generated, including: 'Individuals working under a charismatic leader will have higher task performance than will individuals working under a considerate leader' (Howell and Frost 1989: 245).

One hundred and forty-four students volunteered for the experiment. Their course grades were enhanced by 3 per cent for agreeing to participate. They were randomly assigned to work under one of the three types of leadership. The work was a simulated business task. All three leadership approaches were performed by two female actresses. In broad conformity with the hypotheses, subjects working under charismatic leaders scored generally higher in terms of measures of task performance than those working under the other leaders, particularly the considerate leader.

Research in focus 3.5
A quasi-experiment

Since the mid-1980s, a group of researchers has been collecting medical and psychiatric data on a cohort of over 10,000 British civil servants. The first wave of data collection took place between late 1985 and early 1988 and comprised clinical measurement (for example, blood pressure, ECG, cholesterol) and a self-completion questionnaire that generated data on health, stress, and minor psychiatric symptoms. Further measurements of the same group took place in 1989/90 and 1992/3. The decision in the mid-1980s by the then UK government to transfer many of the executive functions of government to executive agencies operating on a more commercial basis than previously afforded the opportunity to examine the health effects of a major organizational change. Ferrie et al. (1998) report the results of their Phase 1 and Phase 3 data. They distinguished between three groups: those experiencing a change; those anticipating they would be affected by the change; and a 'control group' of those unaffected by the change. The authors found significant adverse health effects among those experiencing and anticipating change compared to the control group, although the extent of the effects of the major organizational change (or its anticipation) varied markedly between men and women. This study uses a quasi-experimental design, in which a control group is compared to two treatment groups. It bears the hallmarks of a classical experimental design, but there is no random assignment. Subjects were not randomly assigned to the three groups. Whether they were affected (or anticipated being affected by the changes) depended on decisions deriving from government and civil service policy.

study may lack what is often called *mundane realism*, it may nonetheless enjoy *experimental realism* (Aronson and Carlsmith 1968). The latter means that the subjects are very involved in the experiment and take it very seriously.

Quasi-experiments

A number of writers have drawn attention to the possibilities offered by *quasi-experiments*—that is, studies that have certain characteristics of experimental designs but that do not fulfil all the internal validity requirements. A large number of different types of quasi-experiments have been identified (Cook and Campbell 1979), and it is not proposed to cover them here. A particularly interesting form of quasi-experiment occurs in the case of 'natural experiments'. These are 'experiments' in the sense of entailing manipulation of a social setting, but as part of a naturally occurring attempt to alter social arrangements. In such circumstances, it is invariably not possible to assign subjects randomly to experimental and control groups. An example is provided in Research in focus 3.5. The absence of random assignment in the research casts a certain amount of doubt on the study's internal validity, since the groups may not have been equivalent. However, the results of such studies are still compelling, because they are not artificial interventions in social life and because their ecological validity is therefore very strong. Most writers on quasi-experimentation discount natural experiments in which there is no control group or basis for comparison (Cook and Campbell 1979), but occasionally one comes across a single group natural experiment that is particularly striking (see Research in focus 3.6). Experimental designs and more especially quasi-experimental designs have been particularly prominent in evaluation research studies (see Key concept 3.5 and Research in focus 3.7).

Possibly because of the various difficulties with quasi-experiments that have been noted in this section, Grant and Wall (2008) have noted that they are used relatively infrequently in organizational research. However, they also note that there may be ways of addressing some of the concerns regarding internal validity that beset quasi-experiments. For example, they suggest that it may be possible to strengthen causal inferences when it is not possible to assign experimental and control group participants randomly and the researcher has limited or no control over the experimental manipulation. This might be done by seeking out further information that will help to discount some of the rival interpretations of a causal link that arise from the lack of a true experimental design. However, it is unlikely that such a view will find favour among writers who adopt a purist view about the need for experimental designs in order to generate robust causal inferences.

Research in focus 3.6
A natural experiment

The effects of television violence on children is one of the most contested areas of social research and one that frequently causes the media to become especially shrill. St Helena in the South Atlantic provided a fascinating laboratory for the examination of the various claims when television was introduced to the island for the first time in the mid-1990s. The television viewing habits of a large sample of schoolchildren and their behaviour are being monitored and will continue to be monitored for many years to come. The project leader, Tony Charlton, was quoted in *The Times* as saying: 'The argument that watching violent television turns youngsters to violence is not borne out . . . The children have been watching the same amounts of violence, and in many cases the same programmes, as British children. But they have not gone out and copied what they have seen on TV' (Midgley 1998: 5). A report of the findings in *The Times* in April 1998 found that 'the shared experience of watching television made them less likely to tease each other and to fight, and more likely to enjoy books' (Frean 1998: 7). The findings derive from 900 minutes of video footage of children at play during school breaks, diaries kept by around 300 of the children, and ratings by teachers. The reports of the research in academic journals confirm that there was no evidence to suggest that the introduction of television had led to an increase in anti-social behaviour (e.g. Charlton et al. 1998, 1999).

Key concept 3.5
What is evaluation research?

Evaluation research, as its name implies, is concerned with the evaluation of such occurrences as social and organizational programmes or interventions. The essential question that is typically asked by such studies is: has the intervention (for example, a new policy initiative or an organizational change) achieved its anticipated goals? A typical design may have one group that is exposed to the treatment (that is, the new initiative), and a control group that is not. Since it is often neither feasible nor ethical to assign research participants randomly to the two groups, such studies are usually quasi-experimental. The use of the principles of experimental design is fairly entrenched in evaluation research, but other approaches have emerged in recent years. Approaches to evaluation based on qualitative research have emerged. While there are differences of opinion about how qualitative evaluation should be carried out, the different views typically coalesce around a recognition of the importance of an in-depth understanding of the context in which an intervention occurs and the diverse viewpoints of the stakeholders (Greene 1994, 2000).

Pawson and Tilley (1997) advocate an approach that draws on the principles of *critical realism* (see Key concept 2.3) and that sees the outcome of an intervention as the result of generative mechanisms and the contexts of those mechanisms. A focus of the former element entails examining the causal factors that inhibit or promote change when an intervention occurs. Pawson and Tilley's approach is supportive of the use of both quantitative and qualitative research methods. Tilley (2000) outlines an early example of the approach in the context of an evaluation of closed-circuit television (CCTV) in car parks. He observes that there are several mechanisms by which CCTV might deter car crime, such as deterrence of offenders, greater usage of car parks, which in itself produces surveillance, more effective use of security staff, and greater sensitivity among drivers to car security. Examples of contexts are: patterns of usage (such as if the car park is one that fills up and empties during rush-hour periods or one that is in more constant use); blind spots in car parks; and the availability of other sources of car crime for potential offenders. In other words, whether the mechanisms have certain effects is affected by the contexts within which CCTV is installed. The kind of evaluation research advocated by Pawson and Tilley maps these different combinations of mechanism and context in relation to different outcomes.

Research in focus 3.7
A quasi-experimental evaluation

Koeber (2005) reports the findings of a quasi-experiment in which he evaluated the use of multimedia presentations (PowerPoint) and a course website (Blackboard) for teaching introductory sociology at a US university. One group of students acted as the experimental group, in that it was taught using these two forms of presenting learning materials simultaneously; the other group acted as a control group and did not experience the multimedia and website methods. There was no random assignment, but in several respects the two groups were comparable. Therefore, this is not a true experiment, but it has the features of a typical quasi-experiment, in that the researcher tried to make the two treatments as comparable as possible. It is an evaluation study, because the researcher is seeking to evaluate the utility of the two teaching methods. The findings are interesting, in that it was found that there was no significant evidence of a difference in the performance of students (measured by their final grades for the course) between those who experienced the newer methods and those who experienced the more traditional ones. However, those students who were taught with the newer methods tended to perceive the course in more favourable terms, in that they were more likely to perceive various aspects of the course (for example, course design, rapport with students, and the value of the course) in a positive way. Also, the experimental groups were less likely to perceive the course demands as difficult and to view the course workload as high.

Significance of experimental design

As was stated at the outset, the chief reason for introducing the experiment as a research design is because it is frequently considered to be a yardstick against which quantitative research is judged. This occurs largely because of the fact that a true experiment will allow doubts about internal validity to be allayed and reflects the considerable emphasis placed on the determination of causality in quantitative research. As we will see in the next section, **cross-sectional designs** of the kind associated with survey research are frequently regarded as limited, because of the problems of unambiguously imputing causality when using such designs.

Logic of comparison

However, before exploring such issues, it is important to draw attention to an important general lesson that an examination of experiments teaches us. A central feature of any experiment is the fact that it entails a *comparison*: at the very least it entails a comparison of results obtained by an experimental group with those engendered by a control group. In the case of the Howell and Frost (1989) experiment in Research in focus 3.4, there is no control group: the research entails a comparison of the effects of three different forms of leadership. The advantage of carrying out any kind of comparison like this is that we understand the phenomenon that we are interested in better when we compare it with something else that is similar to it. The case for arguing that charismatic leadership is an effective, performance-enhancing form of leadership is much more persuasive when we view it in relation to other forms of leadership. Thus, while the specific considerations concerning experimental design are typically associated with quantitative research, the potential of comparison in social research represents a more general lesson that transcends matters of both research strategy and research design. In other words, while the experimental design is typically associated with a quantitative research strategy, the specific logic of comparison provides lessons of broad applicability and relevance. This issue is given more specific attention below in relation to the comparative design.

Key concept 3.6
What is a cross-sectional research design?

A cross-sectional design entails the collection of data on *more than one case* (usually quite a lot more than one) and at *a single point in time* in order to collect a body of *quantitative* or *quantifiable data* in connection with two or more variables (usually many more than two), which are then examined to detect *patterns of association*.

Cross-sectional design

The cross-sectional design is often called a survey design, but the idea of the survey is so closely connected in most people's minds with questionnaires and structured interviewing that the more generic-sounding term *cross-sectional design* is preferable. While the research methods associated with surveys are certainly frequently employed within the context of cross-sectional research, so too are many other research methods, including **structured observation**, content analysis, official statistics, and diaries. All these research methods will be covered in later chapters, but in the meantime the basic structure of the cross-sectional design will be outlined.

The cross-sectional design is defined in Key concept 3.6. A number of elements of this definition have been emphasized.

- *More than one case*. Researchers employing a cross-sectional design are interested in variation. That variation can be in respect of people, families, organizations, nation states, or whatever. Variation can be established only when more than one case is being examined. Usually, researchers employing this design will select a lot more than two cases for a variety of reasons: they are more likely to encounter variation in all the variables in which they are interested; they can make finer distinctions between cases; and the requirements of sampling procedure are likely to necessitate larger numbers (see Chapter 8).
- *At a single point in time*. In cross-sectional design research, data on the variables of interest are collected more or less simultaneously. When an individual completes a questionnaire, which may contain fifty or more variables, the answers are supplied at essentially the same time. This contrasts with an experimental design. Thus, in the classical experimental design, someone in the experimental group is pre-tested, then exposed to the experimental treatment, and then post-tested. Days, weeks, months, or even years may separate the different phases. In the case of the Rosenthal and Jacobson (1968) study, eight months separated the pre- and post-testing of the schoolchildren in the study.
- *Quantitative or quantifiable data*. In order to establish variation between cases (and then to examine associations between variables—see the next point), it is necessary to have a systematic and standardized method for gauging variation. One of the most important advantages of quantification is that it provides the researcher with a consistent benchmark. The advantages of quantification and of measurement will be addressed in greater detail in Chapter 7.
- *Patterns of association*. With a cross-sectional design it is possible to examine relationships only between variables. There is no time ordering to the variables, because the data on them are collected more or less simultaneously, and the researcher does not (invariably because he or she cannot) manipulate any of the variables. This creates the problem referred to in Research in focus 3.2 as 'ambiguity about the direction of causal influence'. If the researcher discovers a relationship between two variables, he or she cannot be certain whether this denotes a causal relationship, because the features of an experimental design are not present. All that can be said is the variables are related. This is not to say that it is not possible to draw causal inferences from research based on a cross-sectional design. As will be shown in Chapter 15, there are a number of ways in which the researcher is able to draw certain inferences about causality, but these inferences rarely have the credibility of causal findings deriving from an experimental design. As a result, cross-sectional research invariably lacks the internal validity that is found in most experimental research (see the examples in Research in focus 3.8 and Thinking deeply 3.1).

In this book, the term 'survey' will be reserved for research that employs a cross-sectional research design and in which data are collected by questionnaire or by structured interview (see Key concept 3.7). This will allow me to retain the conventional understanding of what a survey is while recognizing that the cross-sectional research design has a wider relevance—that is, one that is not necessarily associated with the collection of data by questionnaire or by structured interview.

Reliability, replicability, and validity

How does cross-sectional research measure up in terms of the previously outlined criteria for evaluating quantitative research: reliability, replicability, and validity?

- The issues of reliability and measurement validity are primarily matters relating to the quality of the measures that are employed to tap the concepts in which the researcher is interested, rather than matters to do with a research design. In order to address questions of the quality of measures, some of the issues outlined in Chapter 7 would have to be considered.
- Replicability is likely to be present in most cross-sectional research to the degree that the researcher

Research in focus 3.8
Cross-sectional design and internal validity: an example based on the Health and Lifestyles Survey

Blaxter (1990) reports some of the findings of a large-scale cross-sectional study in which data were collected by three methods: a structured interview; physiological data on each respondent carried out by a nurse; and a self-completion questionnaire. Data were collected from a random sample of around 9,000 individuals. At one point Blaxter shows that there is a relationship between whether a person smokes and his or her diet. But how are we to interpret this relationship? Blaxter is quite properly cautious and does not infer any kind of causal relationship between the two. On the basis of the data, we cannot conclude whether smoking causes diet or whether diet causes smoking or whether the association between the two is actually an artefact of a third variable, such as a commitment or indifference to a 'healthy' lifestyle. There is, therefore, an ambiguity about the direction of causal influence.

Thinking deeply 3.1
Direction of causality: is sex good for you?

An article in the *Guardian*'s Health section reviewed evidence about whether sex is good for you. At one point, the author refers to a study of men that seems to suggest that sex does bring health benefits, but she also has to acknowledge the problem of the direction of cause and effect.

> A study of 1,000 men in Caerphilly found that those who had two or more orgasms a week halved their mortality risk compared with those who had orgasms less than once a month. But while the authors concluded that sex seems to have a protective effect on men's health, it is always possible that the association is the other way around—people who are ill are less likely to have sex in the first place. (Houghton 1998: 14)

Key concept 3.7
What is survey research?

Survey research comprises a cross-sectional design in relation to which data are collected predominantly by questionnaire or by structured interview on *more than one case* (usually quite a lot more than one) and at *a single point in time* in order to collect a body of *quantitative* or *quantifiable data* in connection with two or more variables (usually many more than two), which are then examined to detect *patterns of association*.

spells out procedures for: selecting respondents; designing measures of concepts; administering research instruments (such as structured interview or self-completion questionnaire); and analysing data. Most quantitative research based on cross-sectional research designs specifies such procedures to a large degree.

- Internal validity is typically weak. As has just been suggested above, it is difficult to establish causal direction from the resulting data. Cross-sectional research designs produce associations rather than findings from which causal inferences can be unambiguously made. However, procedures for making causal inferences

from cross-sectional data will be referred to in Chapter 15, though most researchers feel that the resulting causal findings rarely have the internal validity of those deriving from experimental designs.

- External validity is strong when, as in the case of research like Blaxter's (1990) study of Health and Lifestyles (see Research in focus 3.8), the sample from which data are collected has been randomly selected. When non-random methods of sampling are employed, external validity becomes questionable. Sampling issues will be specifically addressed in Chapter 8.
- Since much cross-sectional research makes a great deal of use of research instruments, such as self-completion questionnaires and structured **observation schedules**, ecological validity may be jeopardized because the very instruments disrupt the 'natural habitat', as Cicourel (1982) puts it (see quotation on page 48).

Non-manipulable variables

As was noted at the beginning of the section on experimental design, in much if not most social research it is not possible to manipulate the variables in which we are interested. This is why most quantitative social research employs a cross-sectional research design rather than an experimental one. If we wanted internally valid findings in connection with the smoking–diet relationship investigated by Blaxter (1990) (see Research in focus 3.8), we would need to manipulate one of the variables. For example, if we believed that smoking influences diet (perhaps because smoking is an expensive habit, which may affect people's ability to afford certain kinds of food), we might envisage an experiment in which we took the following steps:

- select a random sample of members of the public who do not smoke;
- establish their current dietary habits;
- randomly assign them to one of three experimental treatments: heavy smokers, moderate smokers, and non-smokers (who act as a control group); and
- after a certain amount of time establish their dietary habits.

Such a research design is almost laughable, because practical and ethical considerations are bound to render it unworkable. We would have to turn some people into smokers, and, in view of the evidence of the harmful effects of smoking, this would be profoundly unethical. Also, in view of the evidence about the effects of smoking, it is extremely unlikely that we would find people who would be prepared to allow themselves to be turned into smokers. We might offer incentives for them to become smokers, but that might invalidate any findings about the effects on diet if we believe that economic considerations play an important role in relation to the effects of smoking on diet. This research is essentially unworkable.

Moreover, some of the variables in which social scientists are interested, and which are often viewed as potentially significant independent variables, simply cannot be manipulated, other than by extreme measures. To more or less all intents and purposes, our ethnicity, age, gender, and social backgrounds are 'givens' that are not really amenable to the kind of manipulation that is necessary for a true experimental design. A man might be able to present himself through dress and make-up as a woman to investigate the impact of gender on job opportunities, as Dustin Hoffman's character did in the film *Tootsie*, but it is unlikely that we would find a sufficient number of men to participate in a meaningful experiment to allow such an issue to be investigated (although Thinking deeply 3.2 provides an interesting case of the manipulation of a seemingly **non-manipulable variable**). Moreover, it could be reasonably argued that, even if we could bring this research design to fruition, the researcher would be examining the effects of only the external signs of gender and would be neglecting its more subjective and experiential aspects. Similarly, while the case of a white man presenting himself as a black man in Thinking deeply 3.2 is interesting, it is doubtful

Thinking deeply 3.2
Manipulating a non-manipulable variable: ethnicity

In the 1950s John Howard Griffin (1961) blackened his face and visible parts of his body and travelled around the American South as a person of colour. He behaved appropriately by keeping his eyes averted to show due deference to whites. He was treated as a black man in a number of ways, such as by having to use water fountains designated for 'coloreds'. Griffin's aim was to experience what it was like being a black person in a period and region of racial segregation.

whether a brief sojourn as a person of colour could adequately capture the experience of being black in the American South. Such an experience is formed by many years of personal experience and the knowledge that it will be an ongoing experience. Thus, although the case described in Thinking deeply 3.2 provides an interesting case of manipulating an apparently non-manipulable variable—ethnicity—it is doubtful whether it could meaningfully be applied to an experimental context, not least because it is doubtful whether sufficient numbers of people could be found to endure the discomforts and inconvenience.

On the other hand, the very fact that we can regard certain variables as givens provides us with a clue as to how we can make causal inferences in cross-sectional research. Many of the variables in which we are interested can be *assumed* to be temporally prior to other variables. For example, we can assume that, if we find a relationship between ethnic status and alcohol consumption, that the former is more likely to be the independent variable because it is temporally prior to alcohol consumption. In other words, while we cannot manipulate ethnic status, we can draw causal inferences from cross-sectional data.

Structure of the cross-sectional design

The cross-sectional research design is not easy to depict in terms of the notation previously introduced, but Figure 3.2 captures its main features, except that in this case Obs simply represents an observation made in relation to a variable.

Figure 3.2 implies that a cross-sectional design comprises the collection of data on a series of variables (Obs_1 Obs_2 Obs_3 Obs_4 Obs_5 . . . Obs_n) at a single point in time, T_1. The effect is to create what Marsh (1982) referred to as a 'rectangle' of data that comprises variables Obs_1 to Obs_n and cases $Case_1$ to $Case_n$, as in Figure 3.3. For each case (which may be a person, household, city, nation, etc.) data are available for each of the variables, Obs_1 to Obs_n, all of which will have been collected at T_1. Each cell in the matrix will have data in it.

Cross-sectional design and research strategy

This discussion of the cross-sectional design has placed it firmly in the context of quantitative research. Also, the evaluation of the design has drawn on criteria associated with the quantitative research strategy. It should also be noted, however, that qualitative research often entails a form of cross-sectional design. A fairly typical form of such research is when the researcher employs **unstructured interviewing** or semi-structured interviewing with a number of people. Research in focus 3.9 provides an illustration of such a study.

While emphatically within the qualitative research tradition, the study described in Research in focus 3.9 bears many research design similarities with cross-sectional studies within a quantitative research tradition, like Blaxter (1990). Moreover, it is a very popular mode of qualitative research. The research was not preoccupied with such criteria of quantitative research as internal and external validity, replicability, measurement validity, and so on. In fact, it could be argued that the conversational interview style made the study more ecologically valid than research using more formal instruments of data

Figure 3.2

A cross-sectional design

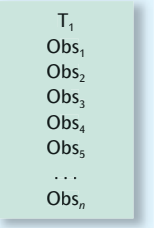


Figure 3.3

The data rectangle in cross-sectional research

	Obs_1	Obs_2	Obs_3	Obs_4	. . .	Obs_n
$Case_1$						
$Case_2$						
$Case_3$						
$Case_4$						
$Case_5$						
. . .						
$Case_n$						

collection. It is also striking that the study was concerned with the factors that influence food selection, like vegetarianism. The very notion of an 'influence' carries a strong connotation of causality, suggesting that qualitative researchers are interested in the investigation of causes and effects, albeit not in the context of the language of variables that so pervades quantitative research. Also, the emphasis was much more on elucidating the *experience* of something like vegetarianism than is often the case with quantitative research. However, the chief point in providing the illustration is that it bears many similarities to the cross-sectional design in quantitative research. It entailed the interviewing of quite a large number of people and at a single point in time. Just as with many quantitative studies using a cross-sectional design, the examination of early influences on people's past and current behaviour is based on their retrospective accounts of factors that influenced them in the past.

Research in focus 3.9
Qualitative research within a cross-sectional design

Beardsworth and Keil (1992) carried out a study of the dietary beliefs and practices of vegetarians. They write that their intention was to contribute 'to the analysis of the cultural and sociological factors which influence patterns of food selection and avoidance. The specific focus is on contemporary vegetarianism, a complex of interrelated beliefs, attitudes and practices . . .' (1992: 253). The authors carried out 'relatively unstructured interviews', which were 'guided by an inventory of issues' with seventy-six vegetarians and vegans in the East Midlands (1992: 261). Respondents were identified through a *snowball sampling* approach. The interviews were taped and transcribed, yielding a large corpus of qualitative data.

Longitudinal design(s)

The longitudinal design represents a distinct form of research design. Because of the time and cost involved, it is a relatively little-used design in social research, so it is not proposed to allocate a great deal of space to it. In the form in which it is typically found in social science subjects such as sociology, social policy, and human geography, it is usually an extension of survey research based on a self-completion questionnaire or structured interview research within a cross-sectional design. Consequently, in terms of reliability, replication, and validity, the longitudinal design is little different from cross-sectional research. However, a longitudinal design can allow some insight into the time order of variables and therefore may be more able to allow causal inferences to be made.

With a longitudinal design a sample is surveyed and is surveyed again on at least one further occasion. It is common to distinguish two types of longitudinal design: the *panel study* and the *cohort study*. With the former type, a sample, often a randomly selected national one, is the focus of data collection on at least two (and often more) occasions. Data may be collected from different types of case within a panel study framework: people, households, organizations, schools, and so on. An illustration of this kind of study is the British Household Panel Survey (BHPS) (see Research in focus 3.10).

In a cohort study, either an entire cohort of people or a random sample of them is selected as the focus of data collection. The cohort is made up of people who share a certain characteristic, such as all being born in the same week or all having a certain experience, such as being unemployed or getting married on a certain day or in the same week. The National Child Development Study (NCDS) is an example of a cohort study (see Research in focus 3.11). A new cohort study—the Economic and Social Research Council (ESRC) Millennium Cohort Study—began at the turn of the present millennium.

Panel and cohort studies share similar features. They have a similar design structure: Figure 3.4 portrays this structure and implies that data are collected in at least two waves on the same variables on the same people. Both panel and cohort studies are concerned with illuminating social change and with improving the understanding of causal influences over time. The latter means that longitudinal designs are somewhat better able to deal with the problem of 'ambiguity about the direction of causal influence' that plagues cross-sectional designs.

Research in focus 3.10
The British Household Panel Survey

The British Household Panel Survey (BHPS) began in 1991, when a national representative sample of 10,264 individuals in 5,538 households were interviewed for the first time in connection with six main areas of interest:

- household organization;
- labour market behaviour;
- income and wealth;
- housing;
- health; and
- socio-economic values.

Panel members are interviewed annually. As a result of the continuous interviewing, it is possible to highlight areas of social change. For example, Laurie and Gershuny (2000) show that there have been changes in the ways in which couples manage their money. Over a relatively short five-year period (1991–5), there was a small decline in the proportion of men having a final say in financial decisions and a corresponding small increase in those reporting equal say, although interestingly these trends refer to aggregated replies of partners—around a quarter of partners give different answers about who has the final say!

For further information, see:
www.iser.essex.ac.uk/ulsc/bhps (accessed 17 January 2011).

The BHPS is being gradually replaced by the Understanding Society Survey, which is based on a panel in the region of 40,000 households. See:
www.understandingsociety.org.uk (accessed 17 January 2011).

Research in focus 3.11
The National Child Development Study

The National Child Development Study (NCDS) is based on all 17,000 children born in Great Britain in the week of 3–9 March 1958. The study was initially motivated by a concern over levels of perinatal mortality, but the data collected reflect a much wider range of issues than this focus implies. Data were collected on the children and their families at age 7. In fact, the study was not originally planned as a longitudinal study. The children and their families have been followed up at ages 11, 16, 23, 33, 41–2, 46, and 50–1. Data are collected in relation to a number of areas, including: physical and mental health; family; parenting; occupation and income; and housing and environment.

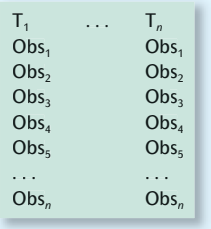
For further information, see Fox and Fogelman (1990); Hodges (1998); and
www.esds.ac.uk/longitudinal/access/ncds/l33004.asp (accessed 17 January 2011).

A new cohort study—the Millennium Cohort Study—began in 2000–1 based on a sample of all children born in England and Wales over a twelve-month period from 1 September 2000 and all children born in Scotland and Northern Ireland from 1 December 2000.

For further information, see:
http://securedata.ukda.ac.uk/sdata/mcs.asp (accessed 17 January 2011).

Figure 3.4

The longitudinal design



Because certain potentially independent variables can be identified at T_1, the researcher is in a better position to infer that purported effects that are identified at T_2 or later have occurred *after* the independent variables. This does not deal with the entire problem about the ambiguity of causal influence, but it at least addresses the problem of knowing which variable came first. In all other respects, the points made above about cross-sectional designs are the same as those for longitudinal designs.

Panel and cohort designs differ in important respects too. A panel study, like the BHPS, that takes place over many years can distinguish between age effects (the impact of the ageing process on individuals) and cohort effects (effects due to being born at a similar time), because its members will have been born at different times. A cohort study, however, can distinguish only ageing effects, since all members of the sample will have been born at more or less the same time. Also, a panel study, especially one that operates at the household level, needs rules to inform how to handle new entrants to households (for example, as a result of marriage or elderly relatives moving in) and exits from households (for example, as a result of marriage break-up or children leaving home).

Panel and cohort studies share similar problems. First, there is the problem of sample attrition through death, moving, and so on, and through subjects choosing to withdraw at later stages of the research. Menard (1991) cites the case of a study of adolescent drug use in the USA in which 55 per cent of subjects were lost over an eight-year period. However, attrition rates are by no means always as high as this. In 1981 the National Child Development Study managed to secure data from 12,537 members of the original 17,414 cohort, which is quite an achievement bearing in mind that twenty-three years would have elapsed since the birth of the children. In 1991 data were elicited from 11,407. The problem with attrition is largely that those who leave the study may differ in some important respects from those who remain, so that the latter do not form a representative group. There is some evidence from panel studies that the problem of attrition declines with time (Berthoud 2000a); in other words, those who do not drop out after the first wave or two of data collection tend to stay on the panel. Secondly, there are few guidelines as to when is the best juncture to conduct further waves of data collection. Thirdly, it is often suggested that many longitudinal studies are poorly thought out and that they result in the collection of large amounts of data with little apparent planning. Fourthly, there is evidence that a *panel conditioning* effect can occur whereby continued participation in a longitudinal study affects how respondents behave. Menard (1991) refers to a study of family caregiving in which 52 per cent of respondents indicated that they responded differently to providing care for relatives as a result of their participation in the research.

Surveys, like the General Household Survey, the British Social Attitudes survey, and the British Crime Survey (see Table 14.1), that are carried out on a regular basis on samples of the population are not truly longitudinal designs because they do not involve the same people being interviewed on each occasion. They are perhaps better thought of as involving a repeated cross-sectional design or trend design in which samples are selected on each of several occasions. They are able to chart change but they cannot address the issue of the direction of cause and effect because the samples are always different.

It is easy to associate longitudinal designs more or less exclusively with quantitative research. However, qualitative research sometimes incorporates elements of a longitudinal design. This is especially noticeable in ethnographic research, when the **ethnographer** is in a location for a lengthy period of time or when interviews are carried out on more than one occasion to address change. See Research in focus 3.12 for an example of the latter.

Most longitudinal studies will be planned from the outset in such a way that sample members can be followed up at a later date. However, it can happen that the idea of conducting a longitudinal study occurs to the researchers only after some time has elapsed. Provided there are good records, it may be possible to follow up sample members for a second wave of data collection or even for further waves. Research in focus 3.13 provides an extremely unusual but fascinating example of a longitudinal design from the USA with both planned and

Research in focus 3.12
Qualitative longitudinal research: the Timescapes project

Qualitative **longitudinal research** (often abbreviated to QLL) that involves repeat qualitative interviews with research participants has become more common since the turn of the century. This is particularly apparent with the 'Timescapes' project, which is a major project that began life in February 2007. The aim is to interview and re-interview people on several occasions to capture social changes and shifts in people's life course and thoughts and feelings. It comprises seven relatively independent projects. Through these projects the researchers aim to track the lives of around 400 people. One of the projects is entitled 'Maculinities, identities and risk: transition in the lives of men as fathers' and aims to get a sense of how masculine identities change in the wake of first-time fatherhood. This particular study builds on research that originally began in Norfolk in 1999, well before the Timescapes project began. Thirty fathers were interviewed in 2000–1 both before and after the birth of their first child. Each man was interviewed three times (two interviews were scheduled after the child's birth). This group of men was then followed up in 2008. A further set of interviews was conducted with eighteen men from south Wales in 2008–9 with the same pattern of one interview before and two interviews after birth. In the course of the interviews use was made of photographs of families and men with their children to stimulate reflection on fatherhood. The use of photographs in interviews is explored in Chapters 19 and 20. The materials will eventually be made available for secondary analysis (see the section on 'Secondary analysis of qualitative data' in Chapter 24).

Sources:

Guardian, 20 Oct. 2009:
www.guardian.co.uk/education/2009/oct/20/timescapes-leeds-research-memories?INTCMP=SRCH

Project website:
www.timescapes.leeds.ac.uk

For information on the masculinities project, see:
www.timescapes.leeds.ac.uk/research-projects/projects/masculinities-fatherhood-risk.php

For some methodological reflections on the Timescapes project, see:
www.timescapes.leeds.ac.uk/methods-ethics

All the above websites were accessed 18 May 2011.

unplanned elements. This is also an interesting illustration of a mixed methods study, in that it combines quantitative and qualitative research.

Case study design

The basic case study entails the detailed and intensive analysis of a single case. As Stake (1995) observes, case study research is concerned with the complexity and particular nature of the case in question. Some of the best-known studies in sociology are based on this kind of design. They include research on:

- a single community, such as Whyte's (1955) study of Cornerville in Boston, Gans's (1962) study of the East End of Boston, M. Stacey's (1960) research on Banbury, and O'Reilly's (2000) research on a community of Britons living on the Costa del Sol in Spain. Increasingly, social researchers are becoming interested in the study of online communities (see Chapter 28 and Research in focus 28.4 in particular);
- a single school, such as studies by Ball (1981) and by Burgess (1983) on Beachside Comprehensive and Bishop McGregor respectively;
- a single family, like O. Lewis's (1961) study of the Sánchez family or Brannen and Nilsen's (2006) investigation of a family of low-skilled British men, which contained four generations in order to uncover changes in 'fathering' over time;

Research in focus 3.13
A planned and unplanned longitudinal design

In the 1940s Sheldon and Eleanor Glueck of the Harvard Law School conducted a study concerned with how criminal careers begin and are maintained. The study entailed a comparison of 500 delinquents and 500 non-delinquents in Massachusetts. The two samples were matched in terms of several characteristics, such as age, ethnicity, and the socio-economic status of the neighbourhoods from which they were drawn. The sample was aged around 14 at the time and was followed up at ages 25 and 33. The data were collected by various means: interviews with the 1,000 participants, their families, and various key figures in their lives (for example, social workers and school teachers); observations of the home; and records of various agencies that had any connection with the participants and their families. Obviously, data concerning criminal activity were collected for each individual by examining records relating to court appearances and parole. While all these sources of data produced quantitative information, qualitative data were also collected through answers to **open questions** in the interviews. Around the mid-1990s Laub and Sampson (2003, 2004) began to follow up the 500 men who had been in the delinquent sample. By this time, they would have been aged 70. Records of death and criminal activity were searched for all 500 men, so that patterns of ongoing criminal activity could be gleaned. Further, they managed to find and then interview fifty-two of the original delinquent sample. These cases were selected on the basis of their patterns of offending over the years, as indicated by the criminal records. The interviews were **life history interviews** to uncover key turning points in their lives and to find out about their experiences. This is an extremely unusual example of a longitudinal study that contains planned elements (the original wave of data collection, followed by the ones eleven and eighteen years later) and an unplanned element conducted by Laub and Sampson many years later.

Research in focus 3.14
A case study

Holdaway (1982, 1983) was a police officer who was also conducting doctoral research on his own police service, which was located in a city. His main research method was **ethnography**, whereby he was a participant observer who observed interaction, listened to conversations, examined documents, and wrote up his impressions and experiences in field notes. Holdaway's superiors did not know that he was conducting research on his own force, so that he was a covert researcher. This is a controversial method on ethical grounds (see Chapters 6 and 19). Holdaway's research provides insights into the nature of police work and the occupational culture with which officers surround themselves.

- a single organization, such as studies of a factory by writers such as Burawoy (1979), and Cavendish (1982), or of pilferage in a single location like a bakery (Ditton 1977), of a single police service (Holdaway 1982, 1983; see Research in focus 3.14), or of a single call centre (Callaghan and Thompson 2002; Nyberg 2009);
- a person, like the famous study of Stanley, the 'jack-roller' (Shaw 1930); such studies are often characterized as using the life history or biographical approach (see the section on 'Life history and oral history interviewing' in Chapter 20); and
- a single event, such as the Cuban Missile Crisis (Allison 1971), the events surrounding the media reporting of a specific issue area (Deacon, Fenton, and Bryman 1999), the Balinese cockfight (Geertz 1973*b*), and the study of a disaster incident (Vaughan 1996, 2004).

What is a case?

The most common use of the term 'case' associates the case study with a location, such as a community or organization. The emphasis tends to be upon an intensive examination of the setting. There is a tendency to associate case studies with qualitative research, but such an

identification is not appropriate. It is certainly true that exponents of the case study design often favour qualitative methods, such as participant observation and unstructured interviewing, because these methods are viewed as particularly helpful in the generation of an intensive, detailed examination of a case. However, case studies are frequently sites for the employment of *both* quantitative and qualitative research, an approach that will receive attention in Chapter 27. Indeed, in some instances, when an investigation is based exclusively upon quantitative research, it can be difficult to determine whether it is better described as a case study or as a cross-sectional research design. The same point can often be made about case studies based upon qualitative research.

As an illustration of the difficulties of writing about case studies, consider the study described in Thinking deeply 3.3. Ostensibly, it is similar to Beardsworth and Keil's (1992) study of vegetarians, in that it is a piece of qualitative research within a cross-sectional design framework (see Research in focus 3.9). However, it has been described as providing 'case-study evidence' by Davies et al. (1994: 157), presumably on the grounds that the fieldwork was undertaken in a single location. I would prefer to reserve the term 'case study' for those instances where the 'case' is the focus of interest in its own right. The study in Thinking deeply 3.3 is no more a case study of Kidderminster than Beardsworth and Keil's (1992) research is based on a case study of the East Midlands. McKee and Bell's (1985) research is concerned with the experience of unemployment among the forty-five couples whom they interviewed. It is not concerned with Kidderminster as such. The town provides a kind of backdrop to the findings rather than a focus of interest in its own right. The crucial point is that Kidderminster is not the unit of analysis; rather it is the sample that is the unit of analysis.

Thinking deeply 3.3
What is the unit of analysis?

McKee and Bell (1985: 387) examined forty-five couples in a single location (Kidderminster in the West Midlands) in order to examine 'the impact of male unemployment on family and marital relations'. They describe their research instrument as an 'unstructured, conversational interview style'. In most cases, husbands and wives were interviewed jointly. The interviews were very non-directive, allowing the couples considerable freedom to answer in their own terms and time. Their research focused on the range of problems faced by unemployed families, the processes by which they cope, and the variations in their experiences. Thus the focus was very much on the experience of unemployment from the perspective of the couples. The authors show, for example, that the impact of husbands' unemployment on their wives is often far greater than is usually appreciated, since research frequently takes the unemployed person as the main hub of the enquiry. Couples often reported changes to the domestic division of labour, which in turn raised questions for them about images of masculinity and identity.

Is this study a case study of unemployment in Kidderminster or is it better thought of as a cross-sectional design study of unemployed men and their wives? As I suggest in the text, it is not terribly helpful to think of it as a case study, because Kidderminster is not the unit of analysis. It is about the responses to unemployment among a sample of individuals; the fact that the interviewees were located in Kidderminster is not significant to the research findings. However, it is not always easy to distinguish whether an investigation is of one kind rather than another. As these reflections imply, it is important to be clear in your own mind what your unit of analysis is.

Similarly, Powell and Butterfield (1997) present a quantitative analysis of promotion decisions in a US government department. They were concerned to investigate how far race had an impact on promotions within the department. The researchers found that race did not have a direct effect on promotion, but it did have an indirect effect. This occurred because race had an impact on two variables—whether the applicant was employed in the hiring department and the number of years of work experience—which in turn affected promotion. The impact of race on these two variables was such that people of colour were disadvantaged with respect to promotion. Once again, we see here a study that has the hallmarks of both a cross-sectional design and a case study, but this time the research strategy was a quantitative one. As with the McKee and Bell (1985) research, it seems better to describe it as employing a cross-sectional design rather than a case study, because the case itself is not the

apparent object of interest: it is little more than a location that forms a backdrop to the findings.

Similarly, I would tend to argue that the study of redundant steelworkers by Westergaard et al. (1989) is a case study of the effects of redundancy in which a quantitative research strategy was employed with clear indications of a cross-sectional design. With a case study, the case is an object of interest in its own right, and the researcher aims to provide an in-depth elucidation of it. Unless a distinction of this or some other kind is drawn, it becomes impossible to distinguish the case study as a special research design, because almost any kind of research can be construed as a case study: research based on a national, random sample of the population of Great Britain would have to be considered a case study of Great Britain! However, it also needs to be appreciated that, when specific research illustrations are examined, they can exhibit features of more than one research design. What distinguishes a case study is that the researcher is usually concerned to elucidate the unique features of the case. This is known as an *idiographic* approach. Research designs like the cross-sectional design are known as *nomothetic*, in that they are concerned with generating statements that apply regardless of time and place. However, an investigation may have elements of both (see Research in focus 3.15).

Research in focus 3.15
A cross-sectional design with case study elements

Sometimes, an investigation may have both cross-sectional and case study elements. For example, Leonard (2004) was interested in the utility of the notion of social capital for research into neighbourhood formation. As such, she was interested in similar issues to the study in Research in focus 2.2. She conducted her study in a Catholic housing estate in West Belfast, where she carried out semi-structured interviews with 246 individuals living in 150 households. Her findings relate to the relevance of the concept of social capital, so that the research design looks like a cross-sectional one. However, on certain occasions she draws attention to the uniqueness of Belfast with its history in recent times of conflict and the search for political solutions to the problems there. At one point she writes: 'In West Belfast, as the peace process develops, political leaders are charged with connecting informal community networks to more formal institutional networks' (Leonard 2004: 939). As this comment implies, it is more or less impossible in a study like this to generate findings concerning community formation without reference to the special characteristics of Belfast and its troubled history.

With experimental and cross-sectional designs, the typical orientation to the relationship between theory and research is a deductive one. The research design and the collection of data are guided by specific research questions that derive from theoretical concerns. However, when a qualitative research strategy is employed within a cross-sectional design, as in Beardsworth and Keil's (1992) research, the approach tends to be inductive. In other words, whether a cross-sectional design is inductive or deductive tends to be affected by whether a quantitative or a qualitative research strategy is employed. The same point can be made of case study research. When the predominant research strategy is qualitative, a case study tends to take an inductive approach to the relationship between theory and research; if a predominantly quantitative strategy is taken, it tends to be deductive.

Reliability, replicability, and validity

The question of how well the case study fares in the context of the research design criteria cited early on in this chapter—measurement validity, internal validity, external validity, ecological validity, reliability, and replicability—depends in large part on how far the researcher feels that these are appropriate for the evaluation of case study research. Some writers on case study research, like Yin (2009), consider that they are appropriate criteria and suggest ways in which case study research can be developed to enhance its ability to meet the criteria; for others, like Stake (1995), they are barely mentioned, if at all. Writers on case study research whose point of orientation lies primarily with a qualitative research strategy tend to play down or ignore the salience of these factors, whereas those writers who have been strongly influenced by the quantitative research strategy tend to depict them as more significant.

However, one question on which a great deal of discussion has centred concerns the *external validity* or *generalizability* of case study research. How can a single case possibly be representative so that it might yield findings that can be applied more generally to other cases? For

example, how could the findings from Holdaway's (1982, 1983) research, referred to in Research in focus 3.14, be generalizable to all police services in Great Britain? The answer, of course, is that they cannot. It is important to appreciate that case study researchers do not delude themselves that it is possible to identify typical cases that can be used to represent a certain class of objects, whether it is factories, mass-media reporting, police services, or communities. In other words, they do *not* think that a case study is a sample of one.

Types of case

Following on from the issue of external validity, it is useful to consider a distinction between different types of case that is sometimes made by writers. Yin (2009) distinguishes five types.

- The *critical case*. Here the researcher has a well-developed theory, and a case is chosen on the grounds that it will allow a better understanding of the circumstances in which the hypothesis will and will not hold. The study by Festinger et al. (1956) of a religious cult whose members believed that the end of the world was about to happen is an example. The fact that the event did not happen by the appointed day allowed the researchers to test the authors' propositions about how people respond to thwarted expectations.
- The *extreme* or *unique case*. The unique or extreme case is, as Yin observes, a common focus in clinical studies. Margaret Mead's (1928) well-known study of growing up in Samoa seems to have been motivated by her belief that the country represented a unique case. She argued that, unlike most other societies, Samoan youth do not suffer a period of anxiety and stress in adolescence. The factors associated with this relatively trouble-free period in their lives were of interest to her, since they might contain lessons for Western youth. Fielding (1982) conducted research on the extreme right-wing organization the National Front. While the National Front was not unique on the British political scene, it was extremely prominent at the time of his research and was beginning to become an electoral force. As such, it held an intrinsic interest that made it essentially unique.
- The *representative* or *typical case*. I prefer to call this an *exemplifying* case, because notions of representativeness and typicality can sometimes lead to confusion. With this kind of case, 'the objective is to capture the circumstances and conditions of an everyday or commonplace situation' (Yin 2009: 48). Thus a case may be chosen because it exemplifies a broader category of which it is a member. The notion of exemplification implies that cases are often chosen not because they are extreme or unusual in some way but because either they epitomize a broader category of cases or they will provide a suitable context for certain research questions to be answered. An illustration of the first kind of situation is Lynd and Lynd's (1929, 1937) classic community study of Muncie, Indiana, in the USA, which they dubbed 'Middletown' precisely because it seemed to typify American life at the time. The second rationale for selecting exemplifying cases is that they allow the researcher to examine key social processes. For example, a researcher may seek access to an organization because it is known to have implemented a new technology and he or she wants to know what the impact of that new technology has been. The researcher may have been influenced by various theories about the relationship between technology and work and by the considerable research literature on the topic, and as a result seeks to examine the implications of some of these theoretical and empirical deliberations in a particular research site. The case merely provides an apt context for the working-through of these research questions. To take a concrete example, Russell and Tyler's (2002) study of one store in the 'Girl Heaven' UK chain of retail stores for 3–13-year-old girls does not appear to have been motivated by the store being critical, unique, or by it providing a context that had never before been studied, but was to do with the capacity of the research site to illuminate the links between gender and consumption and the commodification of childhood in modern society.
- The *revelatory case*. The basis for the revelatory case exists 'when an investigator has an opportunity to observe and analyse a phenomenon previously inaccessible to scientific investigation' (Yin 2009: 48). As examples, Yin cites Whyte's (1955) study of Cornerville, and Liebow's (1967) research on unemployed blacks.
- The *longitudinal case*. Yin suggests that a case may be chosen because it affords the opportunity to be investigated at two or more junctures. However, many case studies comprise a longitudinal element, so that it is more likely that a case will be chosen both because it is appropriate to the research questions on one of the other four grounds and also because it can be studied over time.

Any case study can involve a combination of these elements, which can best be viewed as rationales for choosing particular cases. For example, Margaret Mead's (1928) classic study of growing up in Samoa has been depicted above as an extreme case, but it also has elements of a critical case because she felt that it had the potential to demonstrate that young people's responses to entering their teenage years are not determined by nature alone. Instead, she used growing up in Samoa as a critical case to demonstrate that culture has an important role in the development of humans, thus enabling her to cast doubt on notions of biological determinism.

It may be that it is only at a very late stage that the singularity and significance of the case becomes apparent (Radley and Chamberlain 2001). Flyvbjerg (2003) provides an example of this. He shows how he undertook a study of urban politics and planning in Aalborg in Denmark, thinking it was a critical case. After conducting his fieldwork for a while, he found that it was in fact an extreme case. He writes as follows:

> Initially, I conceived of Aalborg as a 'most likely' critical case in the following manner: if rationality and urban planning were weak in the face of power in Aalborg, then, most likely, they would be weak anywhere, at least in Denmark, because in Aalborg the rational paradigm of planning stood stronger than anywhere else. Eventually, I realized that this logic was flawed, because my research [on] local relations of power showed that one of the most influential 'faces of power' in Aalborg, the Chamber of Industry and Commerce, was substantially stronger than their equivalents elsewhere. Therefore, instead of a critical case, unwittingly I ended up with an extreme case in the sense that both rationality and power were unusually strong in Aalborg, and my case study became a study of what happens when strong rationality meets strong power in the area of urban politics and planning. But this selection of Aalborg as an extreme case happened to me, I did not deliberately choose it. (Flyvbjerg 2003: 426)

Thus, we may not always appreciate the nature and significance of a 'case' until we have subjected it to detailed scrutiny.

One of the standard criticisms of the case study is that findings deriving from it cannot be generalized. Exponents of case study research counter suggestions that the evidence they present is limited because it has restricted external validity by arguing that it is not the purpose of this research design to generalize to other cases or to populations beyond the case. This position is very different from that taken by practitioners of survey research. Survey researchers are invariably concerned to be able to generalize their findings to larger populations and frequently use **random sampling** to enhance the representativeness of the samples on which they conduct their investigations and therefore the external validity of their findings. Case study researchers argue strenuously that this is not the purpose of their craft.

Case study as intensive analysis

Instead, case study researchers tend to argue that they aim to generate an intensive examination of a single case, in relation to which they then engage in a theoretical analysis. The central issue of concern is the quality of the theoretical reasoning in which the case study researcher engages. How well do the data support the theoretical arguments that are generated? Is the theoretical analysis incisive? For example, does it demonstrate connections between different conceptual ideas that are developed out of the data? The crucial question is not whether the findings can be generalized to a wider universe but how well the researcher generates theory out of the findings. This view of generalization is called 'analytic generalization' by Yin (2009) and 'theoretical generalization' by J. C. Mitchell (1983). Such a view places case study research firmly in the inductive tradition of the relationship between theory and research. However, a case study design is not necessarily associated with an inductive approach, as can be seen in the research by Adler and Adler (1985), which was referred to in Chapter 2. Thus, case studies can be associated with both theory generation and theory testing. Further, as M. Williams (2000) has argued, case study researchers are often in a position to generalize by drawing on findings from comparable cases investigated by others. This issue will be returned to in Chapter 18.

Longitudinal research and the case study

Case study research frequently includes a longitudinal element. The researcher is often a participant of an organization or member of a community for many months or years. Alternatively, he or she may conduct interviews with individuals over a lengthy period. Moreover, the researcher may be able to inject an additional longitudinal element by analysing archival information and by retrospective interviewing. Research in focus 3.16 provides an illustration of such research.

Research in focus 3.16
A case study of ICI

Pettigrew (1985) conducted research into the use of organizational development expertise at Imperial Chemical Industries (ICI). The fieldwork was conducted between 1975 and 1983. He carried out 'long semistructured interviews' in 1975–7 and again in 1980–3. During the period of the fieldwork he also had fairly regular contact with members of the organization. He writes: 'The continuous real-time data collection was enriched by retrospective interviewing and archival analysis . . .' (Pettigrew 1985: 40).

Another way in which a longitudinal element occurs is when a case that has been studied is returned to at a later stage. A particularly interesting instance of this is the Middletown study that was mentioned previously. The town was originally studied by Lynd and Lynd in 1924–5 (Lynd and Lynd 1929) and was restudied to discern trends and changes in 1935 (Lynd and Lynd 1937). In 1977 the community was restudied yet again (Bahr et al. 1983), using the same research instruments but with minor changes. Burgess (1987) was similarly concerned with continuity and change at the comprehensive school he had studied in the early 1970s (Burgess 1983) when he returned to study it ten years later. However, as he observes, it is difficult for the researcher to establish how far change is the result of real differences over the two time periods or of other factors, such as different people at the school, different educational issues between the two time periods, and the possible influence of the initial study itself.

Comparative design

It is worth distinguishing one further kind of design: comparative design. Put simply, this design entails studying two contrasting cases using more or less identical methods. It embodies the logic of comparison, in that it implies that we can understand social phenomena better when they are compared in relation to two or more meaningfully contrasting cases or situations. The comparative design may be realized in the context of either quantitative or qualitative research. Within the former, the data-collection strategy will take the form outlined in Figure 3.5. This figure implies that there are at least two cases (which may be organizations, nations, communities, police forces, etc.) and that data are collected from each, usually within a cross-sectional design format.

One of the more obvious forms of such research is in cross-cultural or cross-national research. In a useful definition, Hantrais (1996) has suggested that such research occurs

when individuals or teams set out to examine particular issues or phenomena in two or more countries with the express intention of comparing their manifestations in different socio-cultural settings (institutions, customs, traditions, value systems, life styles, language, thought patterns), using the same research instruments either to carry out secondary analysis of national data or to conduct new empirical work. The aim may be to seek explanations for similarities and differences or to gain a greater awareness and a deeper understanding of social reality in different national contexts.

Figure 3.5
A comparative design

	T_1
	Obs_1
	Obs_2
	Obs_3
Case 1	Obs_4
	Obs_5
	. . .
	Obs_n
	Obs_1
	Obs_2
	Obs_3
Case *n*	Obs_4
	Obs_5
	. . .
	Obs_n

The research by Kelley and De Graaf (1997), referred to in Research in focus 2.4, is an illustration of cross-cultural research that entails a secondary analysis of survey evidence collected in fifteen nations. A further example is Gallie's (1978) survey research on the impact of advanced automation on comparable samples of industrial workers in both England and France. Gallie was able to show that national traditions of industrial relations were more important than technology in explaining worker attitudes and management–worker relations, a finding that was important in terms of the technological determinism thesis that was still current at the time.

Cross-cultural research is not without problems such as: managing and gaining the funding for such research (see Thinking deeply 3.4); ensuring, when existing data, such as official statistics or survey evidence, are submitted to a secondary analysis, that the data are comparable in terms of categories and data-collection methods; ensuring, when new data are being collected, that the need to translate data-collection instruments (for example, interview schedules) does not undermine genuine comparability; and ensuring that samples of respondents or organizations are equivalent. This last problem raises the further difficulty that, even when translation is carried

Thinking deeply 3.4
Forms of cross-cultural research

As its name implies, cross-cultural research entails the collection and/or analysis of data from two or more nations. Possible models for the conduct of cross-cultural research are as follows.

1. A researcher, perhaps in conjunction with a research team, collects data in a number of countries. Gallie's (1978) research on the impact of advanced automation on industrial workers is an illustration of this model, in that he took comparable samples of industrial workers from two oil refineries in both England and France.
2. A central organization coordinates a portion of the work of national organizations. The article by Kelley and De Graaf (1997) that is cited in this chapter provides an example of this model.
3. A secondary analysis is carried out of data that are comparable, but where the coordination of their collection is limited or non-existent. This kind of cross-cultural analysis might occur if researchers seek to ask survey questions in their own country that have been asked in another country. The ensuing data may then be analysed cross-culturally. A further form of this model is through the secondary analysis of officially collected data, such as unemployment statistics. Wall's (1989) analysis of the living arrangements of the elderly in eighteen European countries is an example of such research. The research uncovered considerable diversity in terms of such factors as whether the elderly lived alone and whether they were in institutional care. However, this approach is beset with problems associated with the deficiencies of many forms of official statistics (see Chapter 14) and problems of cross-national variations in official definitions and collection procedures.
4. Teams of researchers in participating nations are recruited by a person or body that coordinates the programme, or alternatively researchers in different countries with common interests make contact and coordinate their investigations. Each researcher or group of researchers has the responsibility of conducting the investigation in his/her/their own country. The work is coordinated in order to ensure comparability of research questions, of survey questions, and of procedures for administering the research instruments (e.g. Crompton and Birkelund 2000). This model differs from (2) above in that it usually entails a specific focus on certain research questions. An example can be found in Research in focus 27.7.
5. Although not genuinely cross-cultural research in the sense of a coordinated project across nations, another form can occur when a researcher compares what is known in one country with new research in another country. For example, Richard Wright, a US criminologist who has carried out a considerable amount of research into street robberies in his own country, was interested in how far findings relating to this crime would be similar in the UK. In particular, US research highlighted the role of street culture in the motivation to engage in such robbery. He was involved in a project that entailed semi-structured interviews with imprisoned street robbers in south-west England (Wright et al. 2006). In fact, the researchers found that street culture played an important role in the UK context in a similar way to that in the USA.

out competently, there is still the potential problem of an insensitivity to specific national and cultural contexts. On the other hand, cross-cultural research helps to reduce the risk of failing to appreciate that social science findings are often, if not invariably, culturally specific. For example, Crompton and Birkelund (2000) conducted research using semi-structured interviewing with comparable samples of male and female bank managers in Norway and Britain. They found that, in spite of more family-friendly policies in Norway, bank managers in both countries struggle to manage career and domestic life. It might have been assumed that countries with greater attachment to such policies would ease these pressures, but comparative, cross-cultural research of this kind shows how easy it is to make such an erroneous inference.

Comparative research should not be treated as solely concerned with comparisons between nations. The logic of comparison can be applied to a variety of situations. The Social Change and Economic Life Initiative, referred to in Research in focus 7.1, entailed identical studies (mainly involving survey research) in six contrasting labour markets, which were chosen to reflect different patterns of economic change in the early to mid-1980s and in the then recent past. By choosing meaningful contrasts, the significance of the different patterns for a variety of experiences of both employers and employees could be portrayed. Such designs are not without problems: the differences that are observed between the contrasting cases may not be due exclusively to the distinguishing features of the cases. Thus, some caution is necessary when explaining contrasts between cases in terms of differences between them.

In terms of issues of reliability, validity, replicability, and generalizability, the comparative study is no different from the cross-sectional design. The comparative design is essentially two or more cross-sectional studies carried out at more or less the same point in time.

The comparative design can also be applied in relation to a qualitative research strategy. When this occurs, it takes the form of a multiple-case study (see Research in focus 3.17). In recent years, a number of writers have argued for a greater use of case study research that entails the investigation of more than one case. Indeed, in certain social science fields, like organization studies, this has become a common research design in its own right. Essentially, a multiple-case (or multi-case) study occurs whenever the number of cases examined exceeds one. The main argument in favour of the multiple-case study is that it improves theory building. By comparing two or more cases, the researcher is in a better position to establish the circumstances in which a theory will or will not hold (Eisenhardt 1989; Yin 2009). Moreover, the comparison may itself suggest concepts that are relevant to an emerging theory.

Related to this point is the fact that there is a growing awareness that the case study and the multiple-case study in particular may play a crucial role in relation to the understanding of causality. However, this awareness reflects a different notion of causality from that outlined earlier in this chapter. In the discussion of independent and dependent variables above, the underlying perception of cause and effect is indicative of what is often referred to as a 'successionist' understanding of causation. As the term 'successionist' implies, the idea of causality entails an effect following on from (that is, succeeding) an independent variable that precedes it. Critical realism (see Key concept 2.3) operates with a different understanding of causation, which is to seek out generative mechanisms that are responsible for observed regularities in the social world and how they operate in particular contexts. Case studies are perceived by writers of a critical realist persuasion to have an important role for research within this tradition, because the intensive nature of most case studies enhances the researcher's sensitivity to the factors that lie behind the operation of observed patterns within a specific context (Ackroyd 2009). The multiple-case study offers an even greater opportunity, because the researcher will be in a position to examine the operation of generative causal mechanisms in contrasting or similar contexts. Thus, Delbridge's (2004) ethnographic study of two 'high-performance' companies in south Wales was able to identify in both firms patterns of resistance and independence that persisted in spite of management efforts to intensify work and to minimize slack in the production process. However, the extent to which informal organization and subversion were found to operate differed considerably between the two firms, and important to this variation was the quality of the relationships between the workers themselves. This represents the causal mechanism producing the variation in resistance between the two factories. The crucial contextual factor was the operation of a blame culture in one of the firms (a Japanese-owned company), whereby any mistake had to be attributed to an individual, which had implications for the quality of relationships among the operatives because of the disputes and disagreements that ensued. Consequently, through the use of a multiple-case study, Delbridge was able to show how variation in informal organization and resistance (an observed regularity) could be understood through its generative causal mechanism (the quality of worker relationships) and through the significance of context (the presence or otherwise of a blame culture).

Research in focus 3.17
A multiple-case study of British companies

In their study of the factors that contribute to competitive success among large British companies, Pettigrew and Whipp (1991) adopted a multiple-case study approach. They examined eight companies, which were made up of a successful and an unsuccessful company in each of three commercial sectors (automobile manufacturing; merchant banking; and book publishing). An additional company drawn from life insurance was also included in the sample. By strategically choosing companies in this way, they could establish the common and differentiating factors that lay behind the successful management of change.

Research in focus 3.17 describes one approach to selecting cases for a multiple-case study. In this illustration, cases were selected on the basis that they represented extreme types—namely, successful and unsuccessful firms, and their operation in certain commercial sectors. Research in focus 3.18 provides another example. In this second example, cases were selected on the basis of quantitative **indicators** of economic deprivation. For example, both the economically deprived areas in Edinburgh and Glasgow were in the top 5 per cent of deprived areas in Scotland. With case selection approaches such as these, the findings that are common to the cases can be just as interesting and important as those that differentiate them. It is also worth pointing out that, although Research in focus 3.17 and 3.18 both used a comparative design using a multiple-case study approach, the former employed a predominantly qualitative research strategy, whereas the latter used a predominantly quantitative one.

Research in focus 3.18
A multiple-case study of Scottish neighbourhoods

Atkinson and Kintrea (2001) were interested in the implications of what are known as *area effects*. Area effects, as their name implies, are to do with the implications of living or working in an area for life chances and attitudes. The issue with which these authors were concerned was to do with the implications of area effects for the experience of poverty among those who are economically deprived. More specifically, is the experience of poverty worse if one lives in a poor area than if one lives in an economically mixed area? Are those who are economically disadvantaged more likely to experience social exclusion in one type of area rather than another (that is, economically deprived or mixed)? The researchers selected an economically disadvantaged area and an economically and socially mixed area in Glasgow for comparison. They selected a similar pair of areas in Edinburgh, thus allowing a further element of comparison because of the greater buoyancy of this city compared to Glasgow. Thus, four areas were selected altogether and samples in each were questioned using a survey instrument. The quantitative comparisons of the data led the researchers to conclude that, by and large, it is 'worse to be poor in a poor area than one which is socially mixed' (Atkinson and Kintrea 2001: 2295).

However, not all writers are convinced about the merits of multiple-case study research. Dyer and Wilkins (1991), for example, argue that a multiple-case study approach tends to mean that the researcher pays less attention to the specific context and more to the ways in which the cases can be contrasted. Moreover, the need to forge comparisons tends to mean that the researcher needs to develop an explicit focus at the outset, whereas critics of the multiple-case study argue that it may be advantageous to adopt a more open-ended approach in many instances. These concerns about retaining contextual insight and a rather more unstructured research approach are very much associated with the goals of the qualitative research strategy (see Chapter 17).

The key to the comparative design is its ability to allow the distinguishing characteristics of two or more cases to act as a springboard for theoretical reflections about contrasting findings. It is something of a hybrid, in that in quantitative research it is frequently an extension of a cross-sectional design and in qualitative research it is frequently an extension of a case study design. It even exhibits certain features that are similar to experiments and quasi-experiments, which also rely on the capacity to forge a comparison.

Bringing research strategy and research design together

Finally, we can bring together the two research strategies covered in Chapter 2 with the research designs outlined in this chapter. Table 3.1 shows the typical form associated with each combination of research strategy and research design and a number of examples that either have been encountered so far or will be covered in later chapters. Table 3.1 refers also to research methods that will be encountered in later chapters but that have not been referred to so far. The Glossary will give you a quick reference to terms used that are not yet familiar to you.

Strictly speaking, Table 3.1 should comprise a third column for mixed methods research, as an approach that combines both quantitative and qualitative research. This has not been done, because the resulting table would be too complicated, since mixed methods research can entail the combined use of different research *designs* (for example, a cross-sectional design and a multiple-case study) as well as methods. However, the quantitative and qualitative components of some of the mixed methods studies referred to in this book *are* included in the table.

Table 3.1

Research strategy and research design

Research design	Research strategy	
	Quantitative	Qualitative
Experimental	*Typical form*. Most researchers using an experimental design employ quantitative comparisons between experimental and control groups with regard to the dependent variable. *Examples*. Research in focus 3.2, 3.4.	No typical form. However, Bryman (1988a: 151–2) notes a study in which qualitative data on schoolchildren were collected within a quasi-experimental research design.
Cross-sectional	*Typical form*. Survey research or structured observation on a sample at a single point in time. Content analysis on a sample of documents. *Examples*. Research in focus 2.9, 3.8, 8.1, 8.4, 12.4, 13.2, 14.1.	*Typical form*. Qualitative interviews or focus groups at a single point in time. Qualitative content analysis of a set of documents relating to a single period. *Examples*. Research in focus 2.3, 2.9, 3.9, 20.4 (see also Table 1.1); Thinking deeply 3.3.
Longitudinal	*Typical form*. Survey research on a sample on more than one occasion, as in panel and cohort studies. Content analysis of documents relating to different time periods. *Examples*. Research in focus 3.10, 3.11, 3.13.	*Typical form*. Ethnographic research over a long period, qualitative interviewing on more than one occasion, or qualitative content analysis of documents relating to different time periods. Such research warrants being dubbed longitudinal when there is a concern to map change. *Examples*. Research in focus 3.12, 17.4.
Case study	*Typical form*. Survey research on a single case with a view to revealing important features about its nature. *Examples*. The choice by Goldthorpe et al. (1968) of Luton as a site for testing the thesis of *embourgeoisement*; the study by Westergaard et al. (1989) of the effects of redundancy at a Sheffield steel plant (Research in focus 7.2).	*Typical form*. The intensive study by ethnography or qualitative interviewing of a single case, which may be an organization, life, family, or community. *Examples*. Research in focus 2.6, 3.14, 19.1, 20.4.
Comparative	*Typical form*. Survey research in which there is a direct comparison between two or more cases, as in cross-cultural research. *Examples*. Research in focus 2.4; Gallie (1978).	*Typical form*. Ethnographic or qualitative interview research on two or more cases. *Examples*. Research in focus 3.17, 3.18, 17.3.

The distinctions are not always perfect. In particular, in some qualitative research it is not obvious whether a study is an example of a longitudinal design or a case study design. Life history studies, research that concentrates on a specific issue over time (e.g. Deacon, Fenton, and Bryman 1999), and ethnography in which the researcher charts change in a single case are examples of studies that cross the two types. Such studies are perhaps better conceptualized as longitudinal case studies rather than as belonging to one category of research design or another. A further point to note is that there is no typical form in the qualitative research strategy/experimental research design cell. Qualitative research in the context of true experiments is very unusual. However, as noted in the table, Bryman (1988*a*) refers to a qualitative study by Hall and Guthrie (1981), which employed a quasi-experimental design.

Key points

- There is an important distinction between a research method and a research design.
- It is necessary to become thoroughly familiar with the meaning of the technical terms used as criteria for evaluating research: reliability; validity; replicability; and the types of validity (measurement, internal, external, ecological).
- It is also necessary to be familiar with the differences between the five major research designs covered: experimental; cross-sectional; longitudinal; case study; and comparative. In this context, it is important to realize that the term 'experiment', which is often used somewhat loosely in everyday speech, has a specific technical meaning.
- There are various potential threats to internal validity in non-experimental research.
- Although the case study is often thought to be a single type of research design, it in fact has several forms. It is also important to be aware of the key issues concerned with the nature of case study evidence in relation to issues like external validity (generalizability).

Questions for review

- In terms of the definitions used in this book, what are the chief differences between each of the following: a research method; a research strategy; and a research design?

Criteria in social research

- What are the differences between reliability and validity and why are these important criteria for the evaluation of social research?
- Outline the meaning of each of the following: measurement validity; internal validity; external validity; and ecological validity.
- Why have some qualitative researchers sought to devise alternative criteria from reliability and validity when assessing the quality of investigations?
- Why have some qualitative researchers *not* sought to devise alternative criteria from reliability and validity when assessing the quality of investigations?

Research designs

- What are the main research designs that have been outlined in this chapter?
- A researcher reasons that people who read broadsheet newspapers are likely to be more knowledgeable about personal finance than readers of tabloid newspapers. He interviews 100 people

about the newspapers they read and their level of financial knowledge. Sixty-five people read tabloids and thirty-five read broadsheets. He finds that the broadsheet readers are on average considerably more knowledgeable about personal finance than tabloid readers. He concludes that reading broadsheets enhances levels of knowledge of personal finance. Assess his reasoning.

Experimental design

- How far do you agree with the view that the main importance of the experimental design for the social researcher is that it represents a model of how to infer causal connections between variables?
- Following on from the previous question, if experimental design is so useful and important, why is it not used more?
- What is a quasi-experiment?

Cross-sectional design

- In what ways does the survey exemplify the cross-sectional research design?
- Assess the degree to which the survey researcher can achieve internally valid findings.
- To what extent is the survey design exclusive to quantitative research?

Longitudinal design(s)

- Why might a longitudinal research design be superior to a cross-sectional one?
- What are the main differences between panel and cohort designs in longitudinal research?

Case study design

- What is a case study?
- Is case study research exclusive to qualitative research?
- What are some of the principles by which cases might be selected?

Comparative design

- What are the chief strengths of a comparative research design?
- Why might comparative research yield important insights?

Online Resource Centre

www.oxfordtextbooks.co.uk/orc/brymansrm4e/

Visit the Online Resource Centre that accompanies this book to enrich your understanding of research designs. Consult web links, test yourself using multiple choice questions, and gain further guidance and inspiration from the Student Researcher's Toolkit.